

Listing of Claims

1-31. (canceled)

32. (currently amended) A graphics subsystem, comprising:

a graphics memory;

a graphics memory access bus connected to said graphics memory;

a plurality of graphics processing units, each graphics processing unit issuing memory requests, each individual memory request having an associated data transfer size; and

a memory controller connected between said graphics memory access bus and said plurality of graphics processing units, said memory controller providing a non-partitioned view of said graphics memory to said plurality of graphics processing units, while dividing said graphics memory access bus into individual bus partitions, each of which is a fraction of the graphics memory access bus size, said memory controller partitioning information within said graphics memory into independently accessible memory partitions, said memory controller routing data ~~from~~ between said independently accessible memory partitions to and said plurality of graphics processing units via said individual bus partitions, said memory controller determining which memory requests are to be serviced in particular clock cycles via one or more of said independently accessible memory partitions and said memory bus partitions;

wherein said memory controller identifies memory requests requiring a subset of said independently accessible memory partitions and determines if another memory request can be serviced in parallel via another subset of said independently accessible memory partitions to improve throughput.

33. (previously presented) The graphics subsystem of claim 32 wherein said memory controller includes control logic to select one or more of said individual bus partitions to route data in response to a data request from a graphics processing unit of said plurality of graphics processing units.

34. (previously presented) The graphics subsystem of claim 32 wherein said memory controller maps data to said independently accessible memory partitions in an interleaved fashion to balance memory load across said independently accessible memory partitions.

35. (previously presented) The graphics subsystem of claims 32 wherein said individual bus partitions have corresponding individual queues.

36. (previously presented) The graphics subsystem of claim 35 further comprising a multiplexer to combine data from said individual queues.

37. (previously presented) The graphics subsystem of claim 35 wherein said individual queues have corresponding arbiter circuits.

38. (previously presented) The graphics subsystem of claim 35 wherein said individual queues facilitate ordered read data delivery and thereby prevent deadlocking across said individual bus partitions.

39. (previously presented) The graphics subsystem of claim 38 wherein said individual queues process read data requests that span a plurality of individual bus partitions.

40. (previously presented) The graphics subsystem of claim 37 wherein said arbiter circuits include an arbiter circuit to prioritize requests from a sub-set of said plurality of graphics processing units.

41. (previously presented) The graphics subsystem of claim 40 wherein said sub-set of said plurality of graphics processing units share a command and write data path.

42. (previously presented) The graphics subsystem of claim 40 wherein each graphics processing unit of said sub-set of said plurality of graphics processing units has a sub-request ID.

43. (previously presented) The graphics subsystem of claim 40 wherein said arbiter circuit pre-arbitrates requests from a sub-set of low-bandwidth graphics processing units.

44. (previously presented) The graphics subsystem of claim 43 wherein said arbiter circuit treats said sub-set of low-bandwidth graphics processing units as a single client.

45. (previously presented) The graphics subsystem of claim 32 wherein individual memory partitions of said independently accessible memory partitions are assigned to solely service individual graphics processing units of said plurality of graphics processing units.

46. (previously presented) The graphics subsystem of claim 32 wherein a selected graphics processing unit of said plurality of graphics processing units accepts data in an out-of-order fashion.

47. (previously presented) The graphics subsystem of claim 46 wherein said selected graphics processing unit accepts data as soon as said data is available from a partition.

48. (previously presented) The graphics subsystem of claim 35 wherein said individual queues include request queues and read data return queues to balance data locality requirements to facilitate memory access efficiency.

49. (previously presented) The graphics subsystem of claim 37 wherein each arbiter circuit implements an independent priority policy to route information to said plurality of graphics processing units.

50. (previously presented) The graphics subsystem of claim 49 wherein said priority policy is a static policy.

51. (previously presented) The graphics subsystem of claim 49 wherein said priority policy is a least recently used policy.

52. (previously presented) The graphics subsystem of claim 49 wherein said priority policy is a round-robin policy.

53. (previously presented) The graphics subsystem of claim 49 wherein said priority policy is a fixed priority policy.

54. (previously presented) The graphics subsystem of claim 49 wherein said priority policy is a dynamic priority policy.

55. (currently amended) A method of servicing data requests from graphics processing units, comprising:

receiving data requests from a plurality of graphics processing units accessing a unitary graphics memory subsystem, each individual data request having an associated data transfer size;

assigning said data requests to ~~one or more~~ a plurality of independently accessible memory partitions imposed upon said unitary graphics memory subsystem; and

delivering data from said independently accessible memory partitions to said graphics processing units via individual bus partitions of a unitary graphics memory access bus;

wherein for a particular request requiring a subset of said independently accessible memory partitions said assigning includes determining if another memory request can be serviced in parallel via a different subset of said independently accessible memory partitions to improve throughput.

56. (previously presented) The method of claim 55 further comprising storing data from said independently accessible memory partitions prior to said delivering.

57. (previously presented) The method of claim 56 further comprising combining stored data prior to said delivering.

58. (previously presented) The method of claim 57 further comprising facilitating ordered read data delivery to prevent deadlocking across said individual bus partitions.

59. (previously presented) The method of claim 58 further comprising prioritizing requests from a sub-set of said plurality of graphics processing units.

60. (previously presented) The method of claim 59 further comprising sharing a command and write data path between a sub-set of said plurality of graphics processing units.

61. (previously presented) The method of claim 60 further comprising assigning a sub-set request ID to individual graphics processing units of said sub-set of said plurality of graphics processing units.

62. (previously presented) The method of claim 55 further comprising pre-arbitrating requests from a sub-set of low-bandwidth graphics processing units of said graphics processing units.

63. (previously presented) The method of claim 62 further comprising treating said sub-set of low-bandwidth graphics processing units as a single client.

64. (previously presented) The method of claim 55 further comprising assigning individual memory partitions of said independently accessible memory partitions to service individual graphics processing units of said graphics processing units.

65. (previously presented) The method of claim 55 further comprising accepting data at a selected graphics processing unit of said graphics processing units in an out-of-order fashion.

66. (previously presented) The method of claim 65 further comprising accepting data as soon as said data is available from a memory partition.

67. (previously presented) The method of claim 55 further comprising balancing data locality requirements to facilitate memory access efficiency.